Is the World Ready for New Memories?

Jim Handy, Objective Analysis

Tom Coughlin, Coughlin Associates

FMS

the Future of Memory and Storage

Why New Memories? Why Now?

- Embedded memories are already changing
  - Embedded NOR is unavailable in advanced process nodes
  - Embedded SRAM has stopped scaling

Let's Have a Look at These

Embedded NOR Replaced in Advanced Process Nodes

- Alternatives available today
  - MRAM, ReRAM

- Multiple foundries already shipping new memories
  - TSMC, Samsung, GLOBALFOUNDRIES, UMC…

- OEMs are starting to embrace these
  - Low power applications
  - High-end MCUs
    - AI is a big driver
  - Slow but steady transition

New Memory Success Stories

Fitness Monitors (Sony, Garmin, Fitbit...)

Smart Glasses

Data Center

Mil/Aero

FMS **Hearing Aids**
the Future of Memory and Storage

Embedded Firmware in Advanced Nodes

- What can be done with MCU & ASIC code?
  - Move it outside the chip, using Serial (SPI) NOR
    - Requires on-chip SRAM caches for code currently being executed
    - The SRAM consumes significantly more area than 28nm NOR would
  - Use chiplet for NOR
    - More on this later
    - Currently an expensive solution
  - Use a non-NOR technology
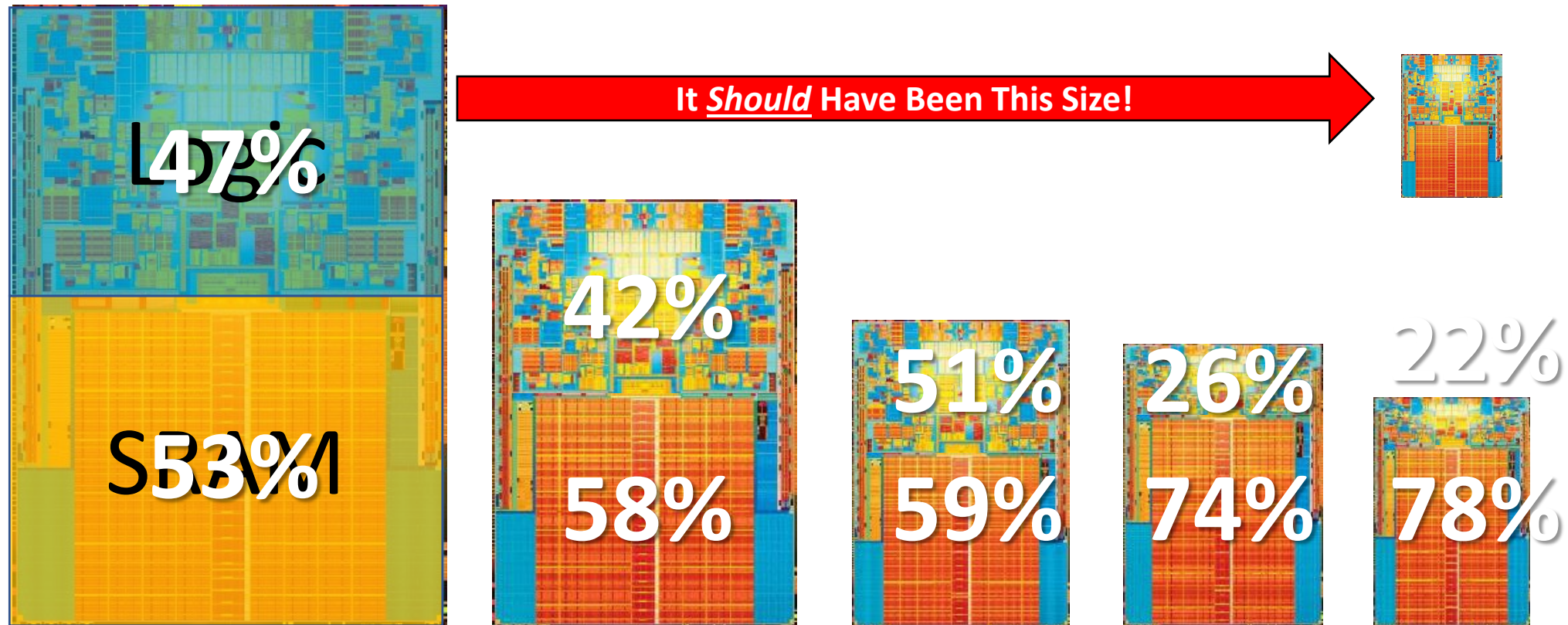   - MRAM, ReRAM are current favorites

SRAM No Longer Scales with Process

SRAM Cell Area vs. Process

Source: Objective Analysis
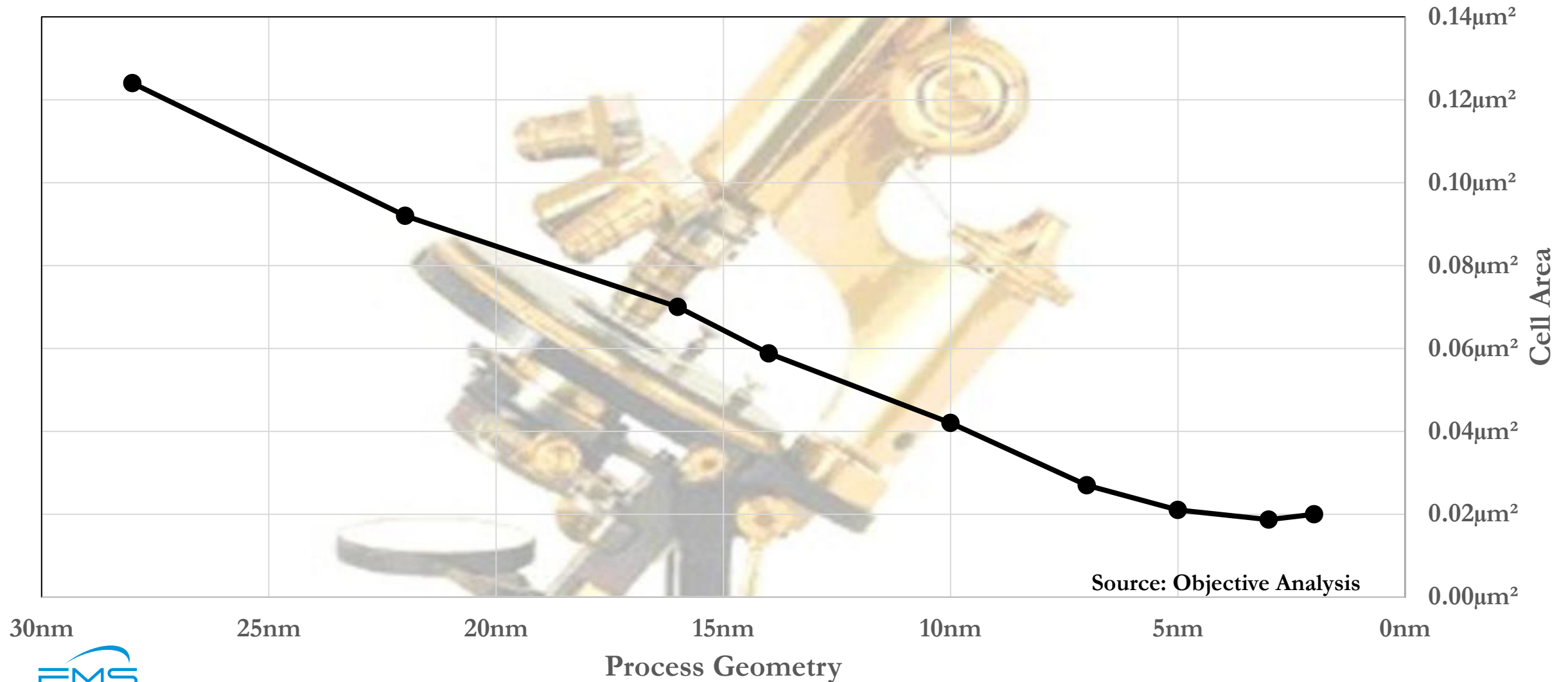
Result: SRAM Caches Barely Shrink

It *Should* Have Been This Size!

Logic 47%

SRAM 53%

42%

58%

51%

59%

26%

74%

22%

78%

Two Interesting Solutions

- Use a new memory technology for cache
  - Smaller cell than SRAM
  - Continues to scale with process
  - Slow and big can substitute for fast and small
- Put the slower caches onto chiplets
  - Cache chiplets on memory wafer
    - New memories are again a good option
  - Logic on foundry logic wafer
- Both of these solutions are nonvolatile/persistent
  - All new memory technologies scale
  - All new memory technologies are persistent

Either Way, Nearly Everything Becomes Persistent

Source: Objective Analysis
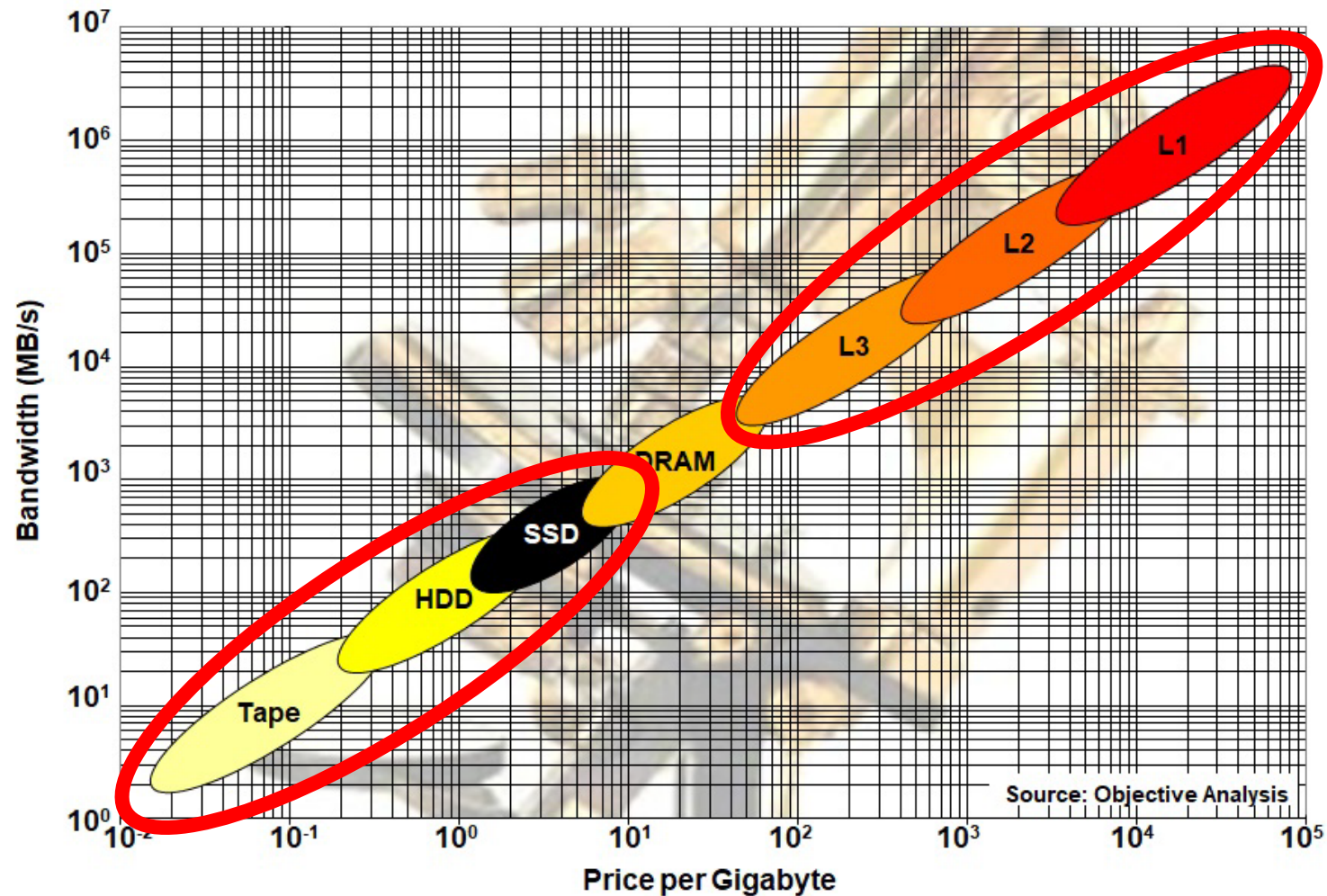
# Why New Memories? Why Now?

- Embedded memories are already changing
    - Embedded NOR doesn't work with FinFET
    - Embedded SRAM has stopped scaling
- DRAM scaling has slowed to a crawl
    - Old way: Each node was ~70% of prior node
    - New way: Past 6 nodes have been 10-20nm
        - 1Xnm, 1Ynm, 1Znm, 1αnm, 1βnm, 1γnm
- All leading memory makers have new memory research projects

Let's Have a Look at These

Enablers Are Being Developed

- SNIA Nonvolatile Memory Programming Model
  - Limited application programs from Oracle & SAP
- CXL supports varying memory speeds
  - Processors can communicate with DRAM, MRAM, ReRAM, etc.
- UCIe is based on CXL

FMS
the Future of Memory and Storage

CXL Supports Any Memory, Volatile or Persistent
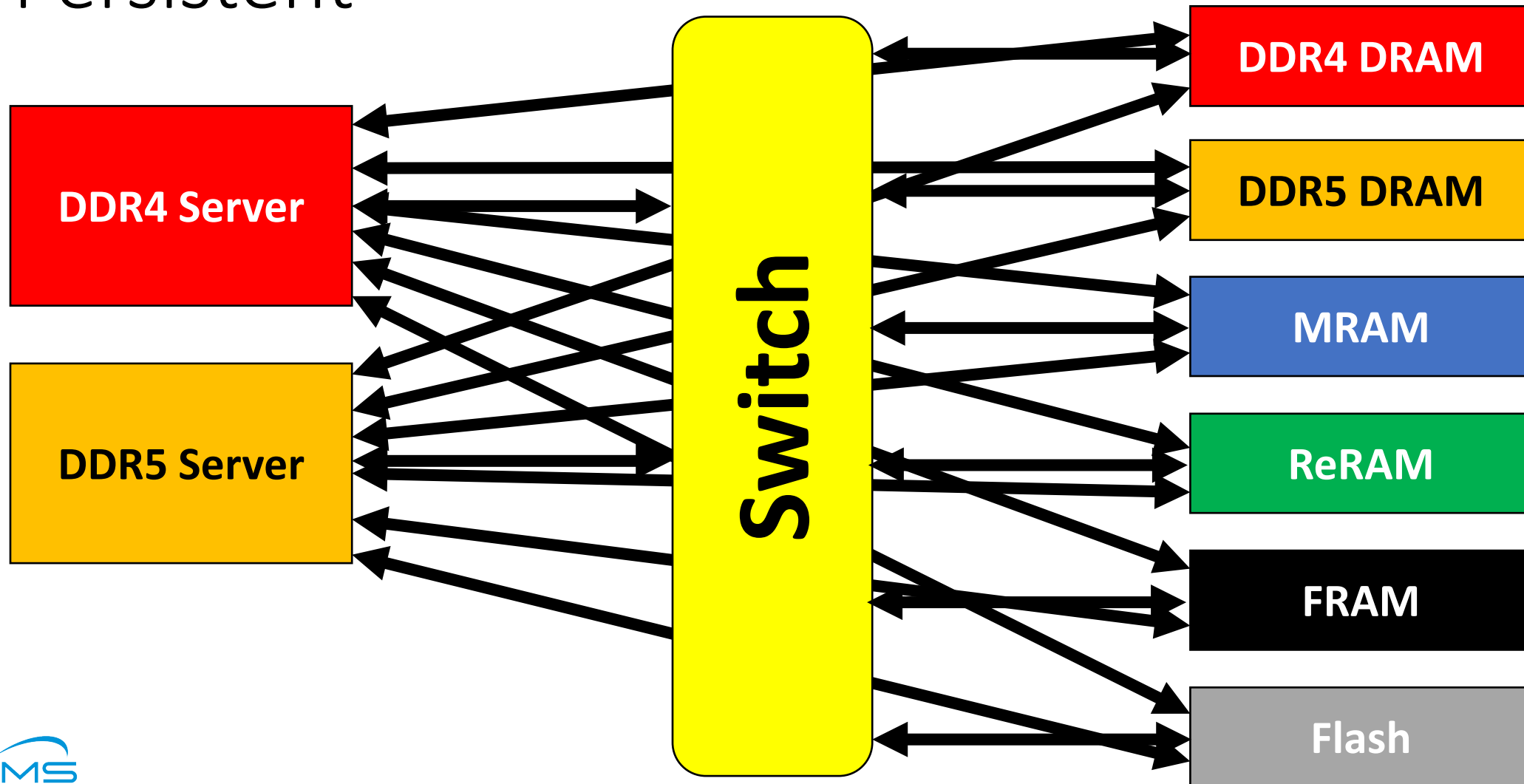
Report: CXL Looks for the Perfect Home

- Covers all perspectives
  - Where CXL is useful, and where it isn't
  - Demand drivers for CXL DRAM modules
  - Opportunities outside of DRAM
  - Forecast (Revenues, units, ASP)
- Available for immediate download:
  - Objective-Analysis.com/reports

What's Been Standing in the Way?

- Economies of Scale determine viability
  - More expensive technologies get relegated to niches
  - Few applications can pay more for better performance

Intel's Optane Proved The Point

- A small die size isn't enough
  - Manufacturing scale determines relative cost
    - Economies of scale prevail
- Optane's volumes were too low
  - Volume never justified the cost
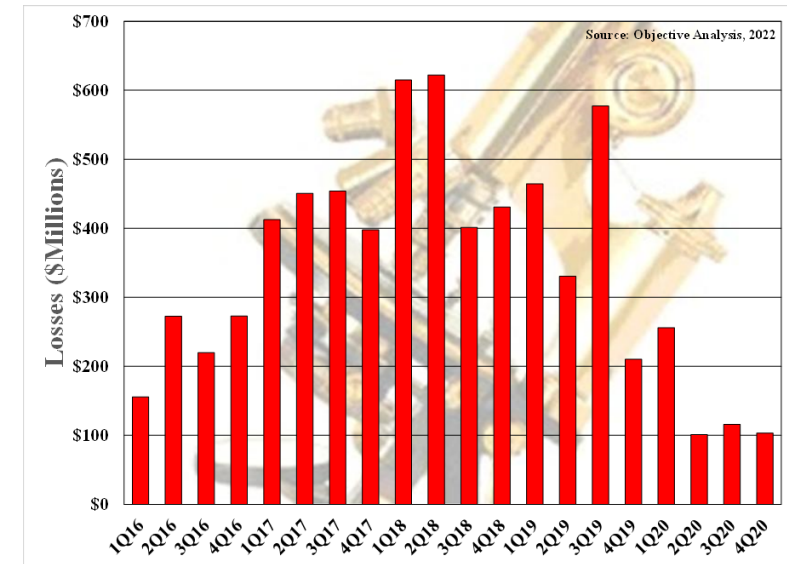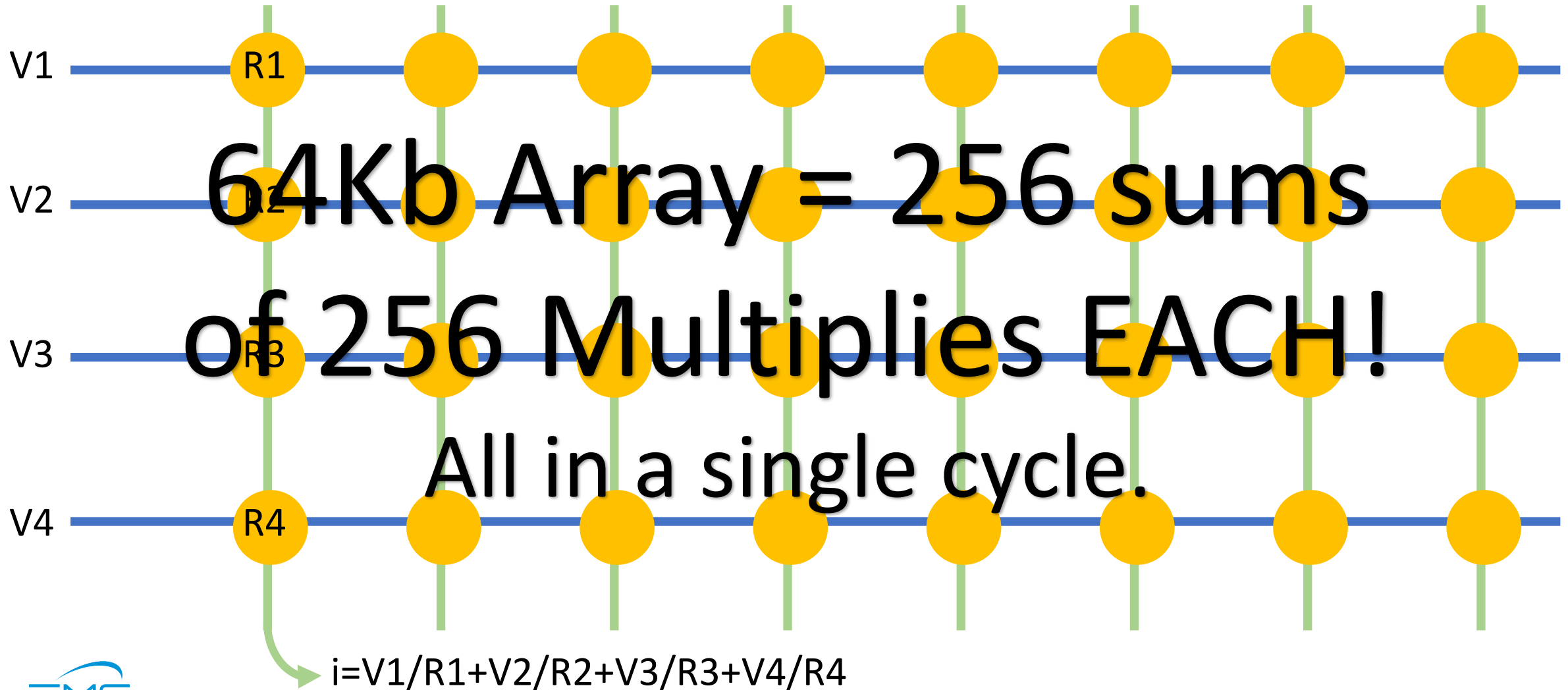 - >$7B in Intel losses
    - Micron losses ~$400M/quarter

**Chart Source:** *Emerging Memories Branch Out*

What About AI?

Neural Networks foredge Applications

Neural Networks Fit New Memory Technologies

64Kb Array = 256 sums of 256 Multiplies EACH!

All in a single cycle.

$i = V1/R1 + V2/R2 + V3/R3 + V4/R4$

# Neural Networks Are Anything <u>But</u> New!

- Intel's 80170NX ETANN
  - Electrically-Trainable Analog Neural Network
- Introduced in 1989
- Not a commercial success
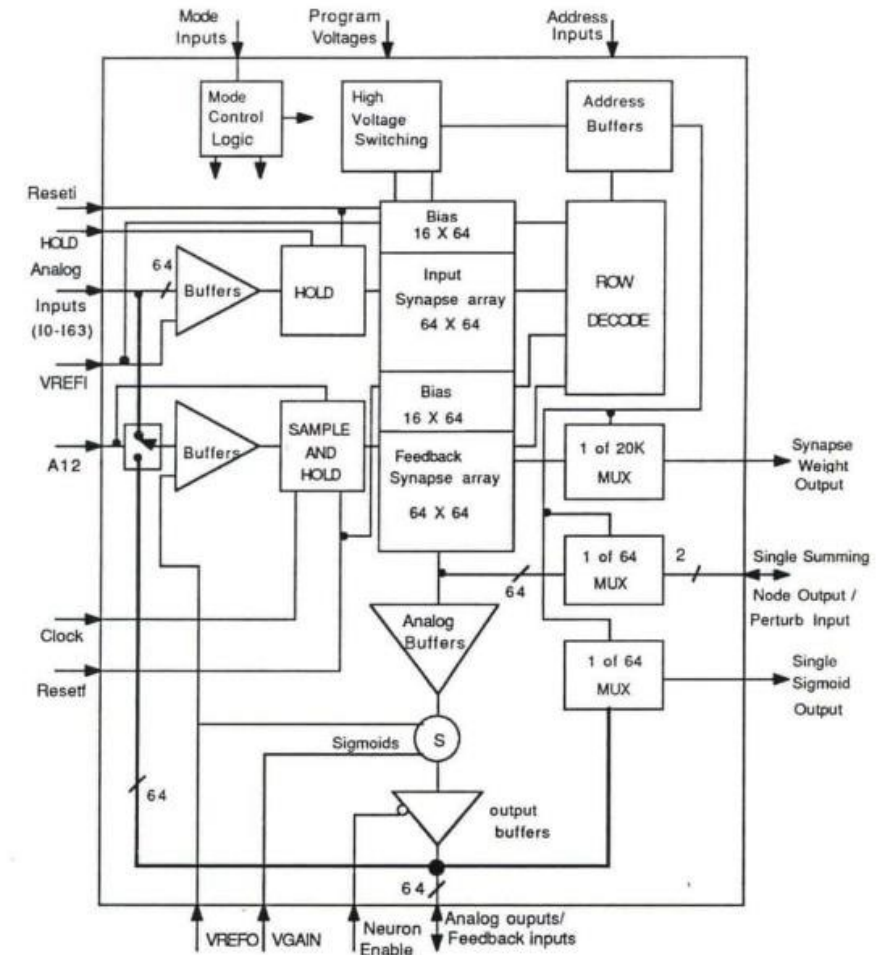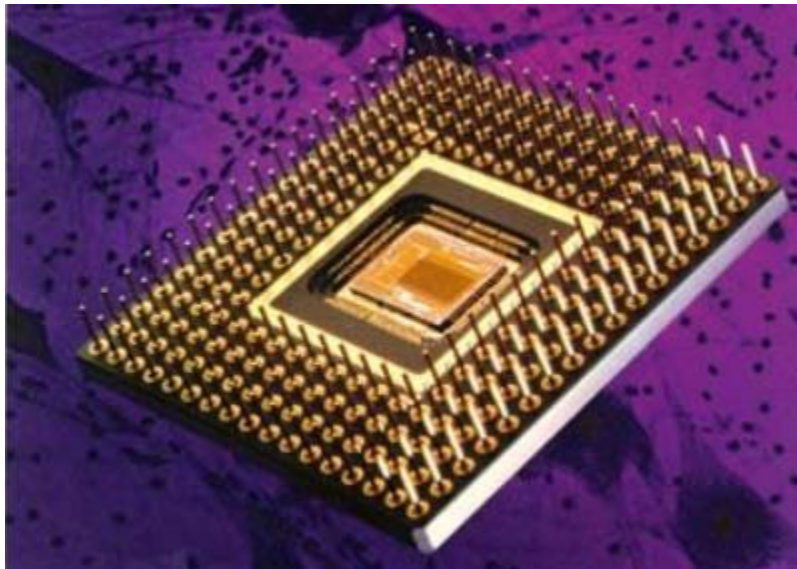
Will Neural Networks Help?

- Neural nets are great for endpoint applications
  - Low energy consumption
  - "Fast enough" performance
  - Low die complexity ► low cost
- There are drawbacks
  - No clear leadership or standards
  - Limited functionality
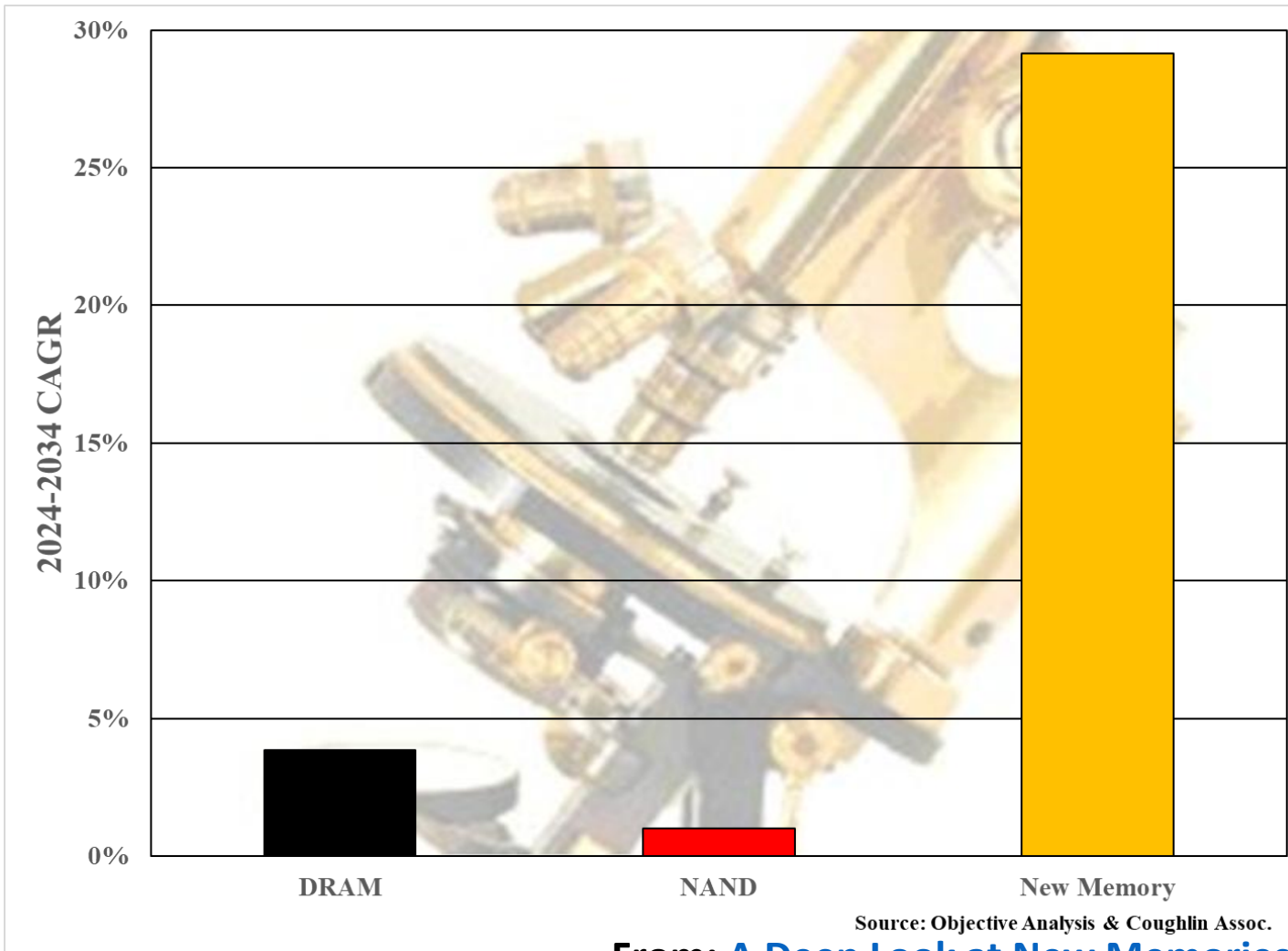 - Very little software support

Neural Nets May Bring AI to the Endpoint

AI

Data Center

AI

AI

AI

Communications Channel

AI

Edge Server

AI

Endpoint
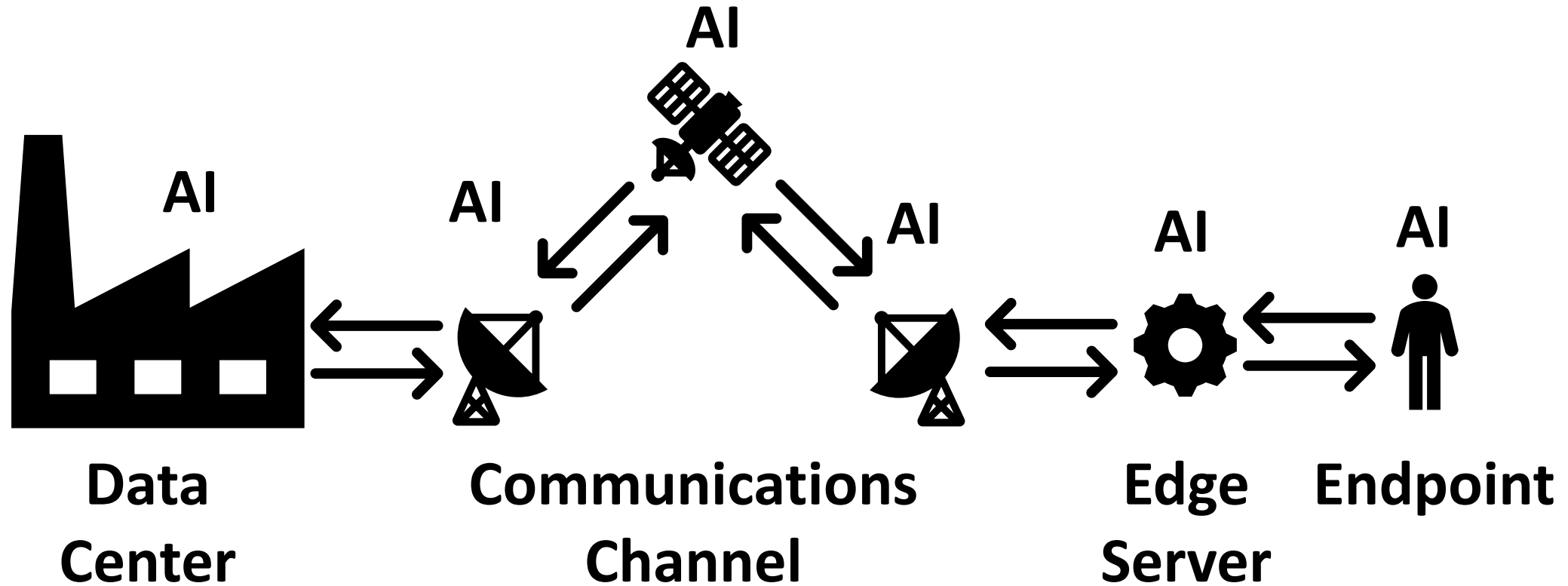
AI eases bandwidth requirements

New Memories Growing Fast

New memory revenue growth much higher than DRAM or NAND flash

Source: Objective Analysis & Coughlin Assoc.

From: A Deep Look at New Memories
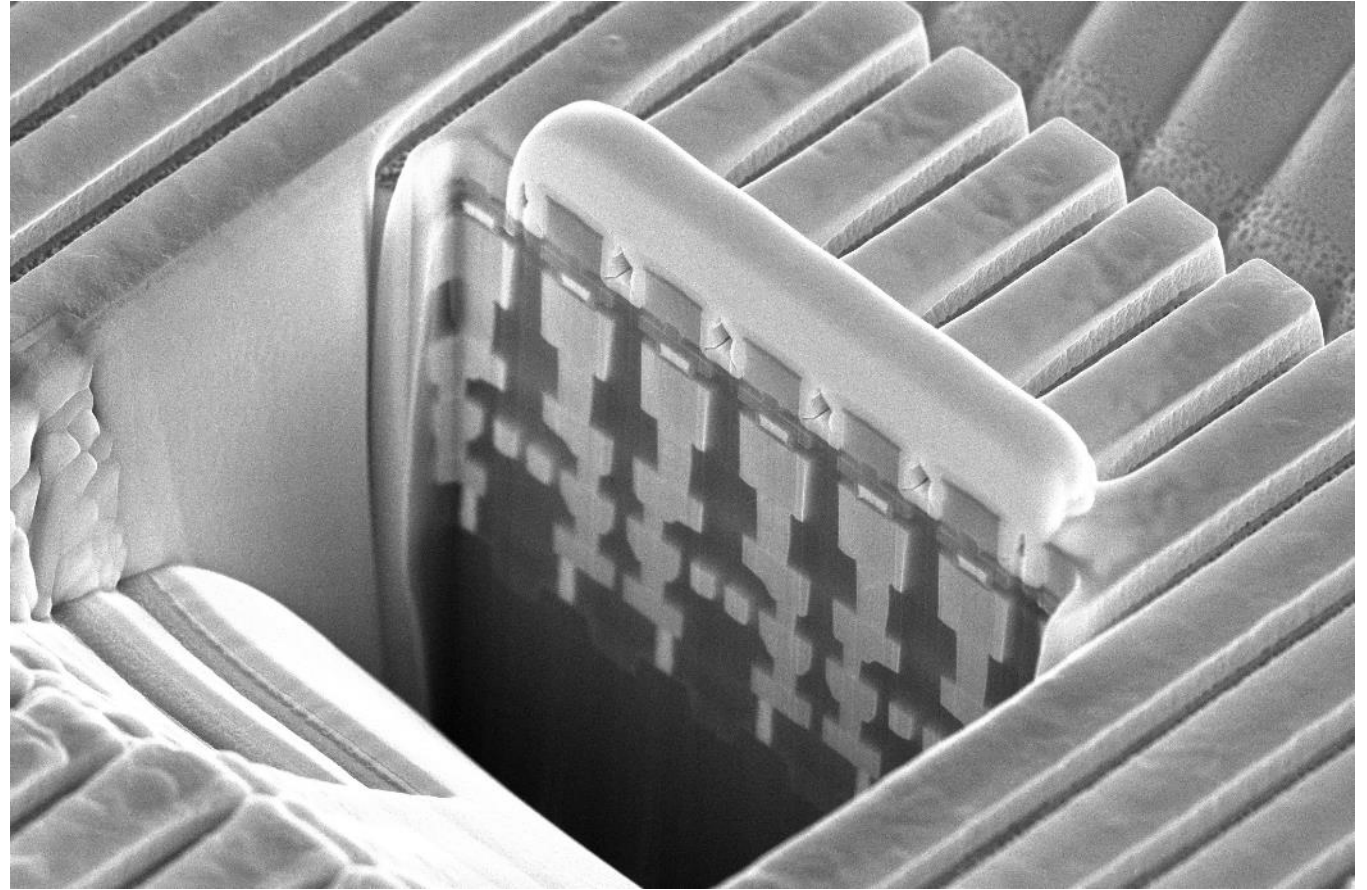
FMS
the Future of Memory and Storage

Report: A Deep Look at New Memories

Photo courtesy of CEA Leti, 2023

Now Available!

https://Objective-Analysis.com/reports/#Emerging
http://www.TomCoughlin.com/techpapers.htm

FMS
the Future of Memory and Storage